
参赛密码 _____
(由组委会填写)

第十二届“中关村青联杯”全国研究生
数学建模竞赛

学 校 上海航天技术研究院（航天八院）

参赛队号 K0127

队员姓名	1.王 涛
	2.王彦革
	3.邱婷婷

参赛密码 _____
(由组委会填写)



第十二届“中关村青联杯”全国研究生 数学建模竞赛

题 目 移动通信中的无线信道“指纹”特征建模

摘 要：

无线信道与周围的环境密切相关，不同环境下的无线信道具有一些差异化的特征。如何发现并提取这些特征并将其应用于优化无线网络，是当前一个研究热点。本文的主要工作是针对无线信道“指纹”特征建模，在先验模型和测试数据的基础上，提取不同场景或不同区域内无线信道的差异化的特征，进而进行分类和归纳。

针对问题一，在对所给数据进行分析后，结合背景资料建立了多径时变信道的复冲激响应模型，以及该模型下的主要参数。本文给出了基于主元数据的特征参数提取方法，即通过主要的特征值与特征向量来代表整个样本数据。由于信道的时变具有相关性，因此可以对大量的样本数据进行去相关性，提取主要的特征值与特征向量，并用这些主要的特征值与特征向量描述信道时变特性。利用基于主元数据的方法，提取的信道参数有多普勒频谱、多普勒带宽、时延扩展等，并与基于统计的方法得出的参数做了对比，提取的参数基本一致，并且由于主元数据较少，因此运算简单。通过基于主元数据提取了信道参数，建立了三维的信道“指纹” $Q = [q_1, q_2, q_3]^T$ ，给出了“指纹”间的“距离” $D = \|Q_1 - Q_2\|$ 以及门限 θ 作为评价指标。在该模型下，三个场景的指纹分别为 $[-499, 157, 1]^T$ ， $[-301, 29, 1]^T$ ， $[-282, 36.4, 1.4]^T$ 通过三维的指纹可以区分出三个场景。

针对问题二，在问题一建立的信道“指纹”模型下，通过基于主元数据的方法提取了两次测试场景的“指纹”，分别为 $[-480, 160, 1]^T$ ， $[-30, 33, 1]^T$ ，通过与问题一中的三个场景对比，得出结论：Test1ForScene 属于场景一，Test2ForScene 属于场景二。

针对问题三，在对所给数据分析的基础上，根据指纹中的延时扩展参数将

测试路段粗略分为 5 段，然后利用问题一中的指纹模型参数提取的方法提取每一路段的指纹，每一路段的指纹差别明显，验证了分段的正确性。所划分的路段为（单位：米）：[0,40]、[40,60]、[60,67]、[67,100]、[100,150]。更为精细划分仍然可以从另外一个维度上进行，但是这会对后续的判决带来更大的误差。

针对问题四，在问题三根据信号时延扩展做出粗略分段的基础上，判断 SampleForTest1 属于问题三中的路段 1, SampleForTest2 属于问题三中的路段 4。为了更为精确的判断两次测量的位置，根据信道的时变特性对路段 1 和路段 4 进行更精细的划分，提取更细致的指纹特征。用两次测量所得的指纹与精细路段划分所得的指纹进行对比，判断 SampleForTest1 可能来自于问题三中路段的 10m—30m 处，SampleForTest2 来自于路段的 75m—90m 处。

关键词：信道“指纹” 主元数据 特征向量 多普勒频谱 延时扩展

目录

一. 问题重述.....	- 5 -
1、问题背景.....	- 5 -
2、问题提出.....	- 7 -
二. 问题假设与符号说明.....	- 7 -
1、问题假设.....	- 7 -
2、符号说明.....	- 8 -
三. 问题一的分析与建模.....	- 8 -
1、问题分析.....	- 8 -
2、信道模型建立.....	- 10 -
3、信道参数的提取.....	- 12 -
4、基于主元数据的特征提取方法:	- 14 -
5、信道“指纹”模型的建立.....	- 19 -
四. 问题二的求解.....	- 20 -
五. 问题三的分析与建模.....	- 22 -
1、问题分析.....	- 22 -
2、基于信道“指纹”的区域划分.....	- 23 -
六. 问题四的求解.....	- 24 -
七. 模型评价与结论.....	- 28 -
参考文献.....	- 29 -

一. 问题重述

1、问题背景

随着科学技术的不断进步和经济水平的逐渐提高，无线通信已成了我们日常生活中不可缺少的必备品，人们对通信质量的要求也越来越高。而整个无线通信中最为关键部分是无线信道，由于发射机与接收机之间的传播路径上分布有复杂的地形地物，其信道往往具有复杂时变的电磁传播特性，如何发现并提取这些特性并将其应用于优化无线网络，是当前一个研究热点。类比人类指纹，我们将上述无线信道的差异化的特征称为无线信道“指纹”。利用无线信道“指纹”特征建模就是在先验模型和测试数据的基础上，提取并利用不同场景或不同区域内无线信道的差异化特征，分析归纳出“指纹”的“数学模型”。真实信道测量结果受多径数目、多径时延、时延扩展、Doppler 扩展等各种参数的影响。

信号在传播过程中会遇到很多建筑物、树木以及起伏的地形，发送和接收之间通常会存在多出一条的信号传播路径，从而会引起能量的吸收和穿透以及电波的反射、散射、绕射等，如下图所示。因此移动信道是充满反射波的传播环境。由于电波通过各个路径的距离不同，因而各路径来的反射波到达时间不同，相位也就不同。不同相位的多个信号在接收端叠加，导致接收信号的幅度发生急剧变化，即产生了衰落。这种衰落是由多径引起的，所以称为多径衰落。

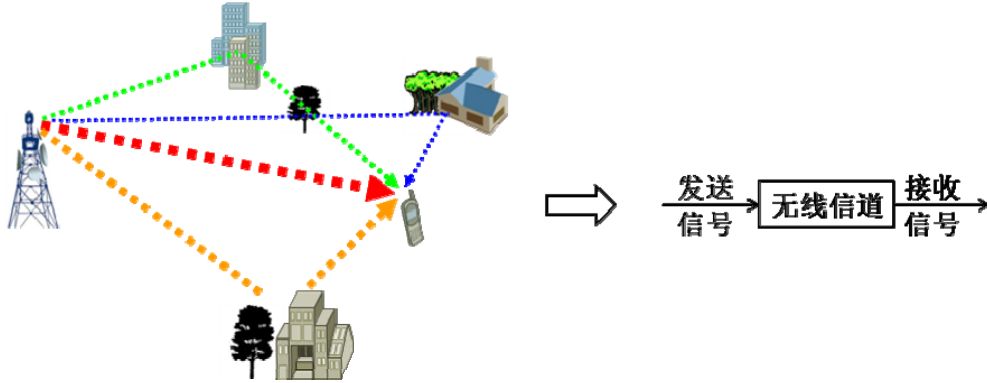


图 1 信号传输模型

在真实无线通信系统中，为了改善信号的传输质量，通常需要在系统的发射端和接收端，各增加一个滤波器。考虑到信道的时变性以及实测中引入了噪声，不同时刻的真实信道测量结果及其对应的无线信道分别为：

$$r[k, n] = \sum_{m=0}^{M-1} h[k-m, n] \cdot g[m] + u[k, n], k=0, 1, \dots, K-1, n=0, 1, \dots, N-1 \quad (1)$$

$$h[k, n] = \sum_{l=0}^{L-1} h_l \cdot \delta[k - \tau_l], k=0, 1, \dots, K-1, K \geq \max_l \{\tau_l\} \quad (2)$$

其中， n 表示测试的样本标识，对应测试时刻，这里假设共有 N 个测试样本； h_l 表示第 n 时刻第 l 条路径上的信道系数； k 表示离散信号的样点标识，这里假设更有 n 个样点； τ_l 表示 n 时刻第 l 条路径上的延迟点数； $r[k, n]$ 表示 n 时

刻的实际接收信号； $h[k,n]$ 表示第 n 时刻的理想信道输出结果； $u[k,n]$ 表示 n 时刻的第 k 个样本引入的复高斯白噪声。需要注意的是，参数 h_l 、 τ_l 和 L 都具有时变性。

由于收发端的相对运动或者信道中反射物体的相对运动会引起多普勒效应，不同来波的多普勒频移不同，引起频率色散，造成信道的时变特征。下图2给出了多普勒效应的产生原理。无线电波从 S 出发移动台在 A 点与 B 点接收信号时的路程差为：

$$\Delta x_i = d \cdot \cos \theta_i = v \cdot \Delta t \cdot \cos \theta_i \quad (3)$$

这里 Δt 是移动台从 A 点运到 B 点所需要的时间； θ_i 是 A 和 B 出接收信号的入射角，由于发射端距离很远，可以近似认为 A 、 B 两处的入射角相等。所以，由于路程差造成的信号相位变化值为：

$$\Delta \varphi = \frac{2\pi d}{\lambda} = \frac{2\pi v \Delta t}{\lambda} \cos \theta_i \quad (4)$$

由此得到频率变化值为：

$$f_d = \frac{1}{2\pi} \cdot \frac{\Delta \varphi}{\Delta t} = \frac{v}{\lambda} \cos \theta_i \quad (5)$$

式中， f_d 称为多普勒频移， λ 为波长， $\frac{v}{\lambda}$ 与入射角无关，表示 f_d 的最大值，称为最大多普勒频移。

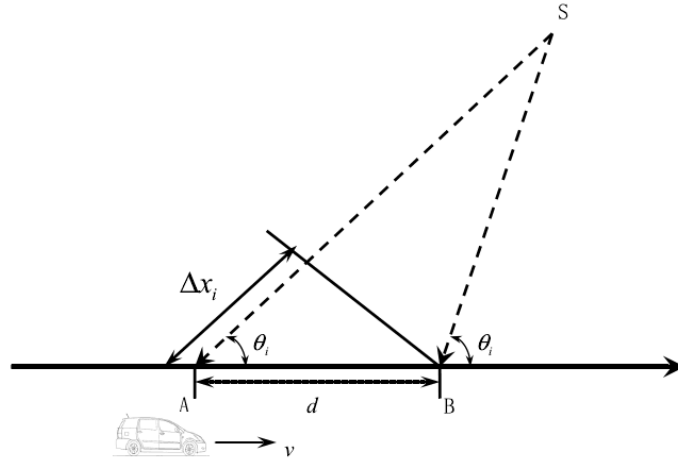


图2 多普勒效应示意图

为了将原有复杂的数据降维,揭示隐藏在复杂数据背后的信息,可以利用主成分分析方法(Principal component analysis)。PCA是一种掌握事物主要矛盾的统计分析方法,它可以从多元事物中解析出主要影响因素,揭示事物的本质,简化复杂的问题,对某些复杂数据进行简化。主要原理是设法将原来变量重新组合成一组新的互相无关的几个综合变量,同时根据实际需要从中可以取出几个较少的总和变量尽可能多地反映原来变量的信息,从而简化数据处理,得出结论。

2、问题提出

问题一：“数据包 1”提供了三种场景的真实信道测量结果。基于三种已知场景的测量结果，参考背景中的知识介绍，采用数学建模特别是特征提取的方法，对所提供的样本加以分析，给出无线信道“指纹”的模型。在此基础上，给出简洁而明确的“评价指标”，用于分析验证所建模“指纹”合理有效。所谓有效是指，该模型应能从数学上对已知的三种场景进行合理区分。要求详细说明建模的思路、使用的方法以及得出结论的过程。另外，所提取的特征或所建立的模型最好有一定的物理意义。最后，提供获取“指纹”的程序代码，要求代码可运行，但代码形式不限。

问题二：“数据包 2”提供了与“问题 1”中某些场景相对应的二个真实信道测量结果。基于“问题 1”中所提供的三个场景，以及所建立的“指纹”模型和“评价指标”，采用数学的方法，识别出此处提供的二个样本分别属于哪个场景。需要保证“场景识别”的结果正确，且对识别的结果进行合理的分析。请明确给出“场景识别”的结果，并详细描述分析的过程。

问题三：“数据包 3”提供了一条连续路段的真实信道测量结果。该结果对应于以 3km/h 的速度步行近 150m 的测试距离。该路段可能包含不同的场景或环境，对应不同的“指纹”特征。基于上述测量结果，先自行分段，采用“问题 1”中的建模方法，给出分段的“指纹”分析。在分段分析的基础上，对所提取的“指纹”进行合理分类。通过对比不同段的分析结果，最终决定该路段可以依“指纹”划分为多少个区域。理论上，“区域划分”越细，后续做“区域识别”的精确程度越高，但过细的“区域划分”会在“指纹”特征中引入更多的错误，导致误判概率增大。要求详细说明“区域划分”的思路及过程，并对划分结果进行合理的分析。

问题四：“数据包 4”提供了二个真实信道测量结果。基于“问题 3”中的“区域划分”和“问题 1”中的“评价指标”，首先判断此处提供的二个样本是否采集自“问题 3”中所提供的路段。对于已判断出的采自上述路段的样本，请识别其对应于“问题 3”中的哪一块区域。需要保证“样本判断”的结果正确，以及“区域识别”的误判距离尽可能小，同时对所识别的结果进行合理的分析。要求详细说明样本判别的思路和方法，以及得出结论的过程。

二. 问题假设与符号说明

1、问题假设

(1) 假设对信道的每一次采样时间内，信道是时不变的；不同采样点间，信道是时变的（即数据包中每一行数据所描述的信道是时不变的，所有行数据之间信道是时变的）；

(2) 假设信道是广义平稳信道；

(3) 假设噪声是高斯的，功率谱在整个频带内是均匀的。

2、符号说明

表 1 符号说明

符号	说明
τ_l	第 l 条路径的时延
$h_l(t)$	t 时刻第 l 条路径的复增益（信道系数）
α_l^2	第 l 条路径的平均功率
B_{coh}	相干带宽
f_c	发射信号的载频
v	接收机与发射机的相对移动速度
$J_0(\cdot)$	一类零阶贝塞尔函数
f_d	最大多普勒频移
T_c	相干时间
ρ_τ	时延扩展系数
Q	信道“指纹”
D	指纹间的“距离”
θ	指纹判别“门限”

三. 问题一的分析与建模

“数据包 1”提供了三种场景的真实信道测量结果。基于三种已知场景的测量结果，参考背景中的知识介绍，采用数学建模特别是特征提取的方法，对所提供的样本加以分析，给出无线信道“指纹”的模型。在此基础上，给出简洁而明确的“评价指标”，用于分析验证所建模“指纹”合理有效。所谓有效是指，该模型应能从数学上对已知的三种场景进行合理区分。

1、问题分析

问题一所要解决的问题是建立适当的模型，以提取信道的参数，典型的信道参数包括信道功率时延分布、均方根时延扩展、相干带宽、信道相关函数与多普勒频谱等。

数据包中测试数据都是冲击信号经过发射滤波器、信道和接收滤波器后得到的，在发射信号和滤波器都固定的条件下，信道信息（滤波器带宽范围内）都体现在接收信号上。考虑到问题中有三个场景，为了直观分析，对每个场景

中的 5 个 Test 分别做等高图如下

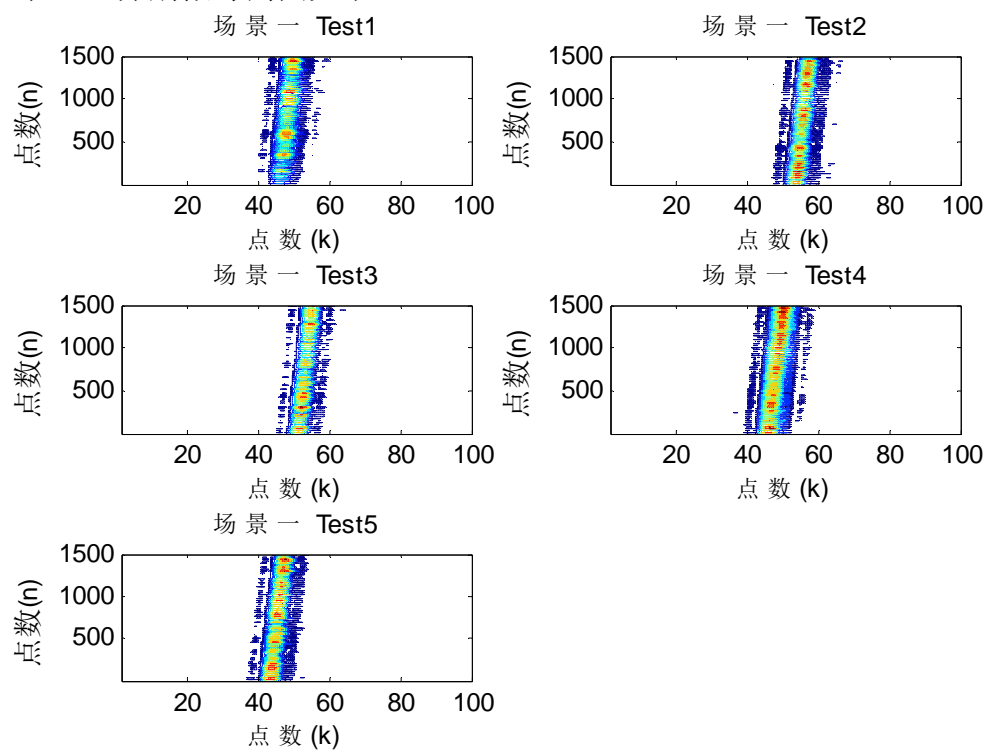


图 3(a) 场景一数据的等高图

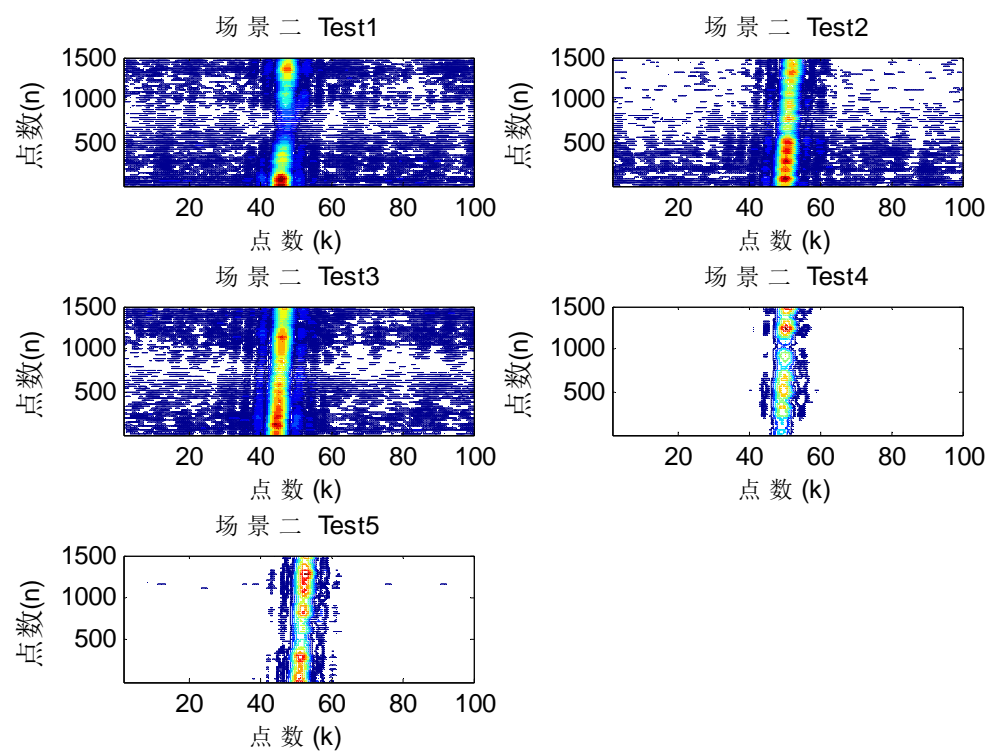


图 3(b) 场景二数据的等高图

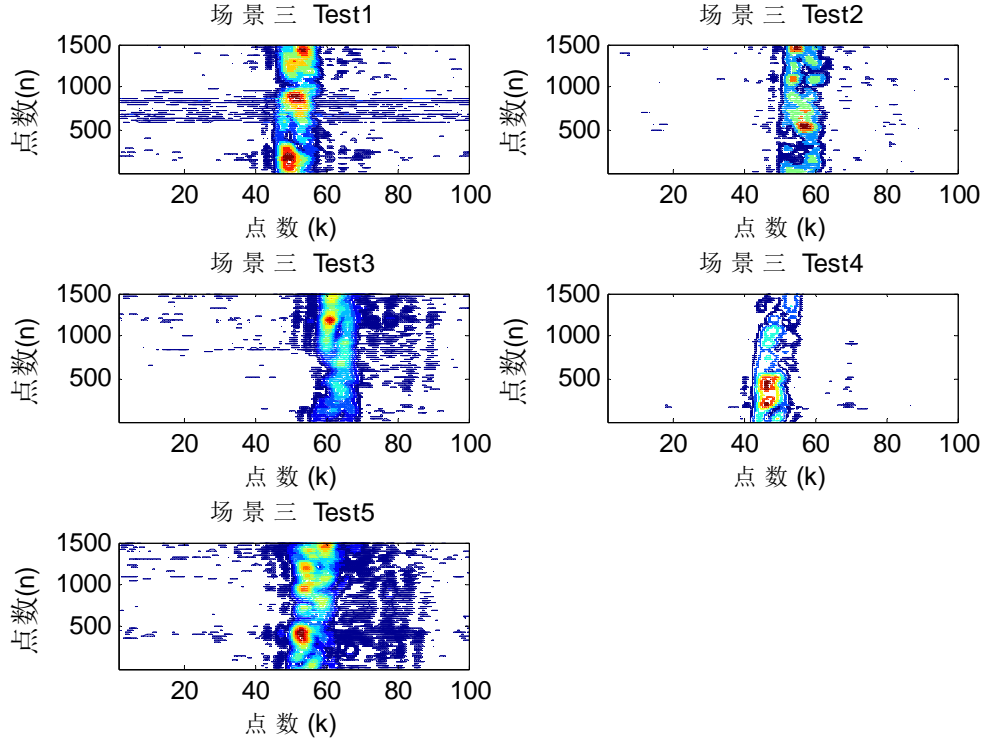


图 3(c) 场景三数据的等高图

从以图中可以看出，三个场景中存在相同点，随着样本采集点的变化出现峰值的时刻也在延后。但是也有很明显的区别：场景一中的峰值出现的时刻随 n 变化较大（体现在倾斜的程度），场景二、三中峰值出现的时刻随 n 变化则相对较小；另外，场景一、二中峰值的宽度（沿 k 方向上的宽度）相对较小，而场景三中峰值的宽度则较宽。

从等高图可以得到如下分析：峰值出现的时刻随时间样本 n 在变化，说明接收设备相对信号发射天线存在移动，那么信号中就会存在多普勒频率，场景一中峰值出现的时刻随 n 变化较快，说明运动速度较大，那么（相对场景二、三）多普勒偏移就会越大，多普勒频谱也会更宽；此外，可以看到三幅图中峰值出现的时刻都随 n 增大延后，说明，接收设备相对发射天线的运动径向速度是背离发射源，多普勒频率应为负值；场景三中峰值的宽度明显比场景一、二较宽，说明场景三中均方根时延扩展更大。

2、信道模型建立

时变多径信道复冲激响应可以描述为

$$h(\tau, t) = \sum_{l=0}^{L-1} h_l(t) \delta(\tau - \tau_l) \quad (6)$$

式中 τ_l 为第 l 条路径的时延， $h_l(t)$ 为 t 时刻第 l 条路径的复增益。

功率时延分布表示为当发送一个 δ 脉冲时，接收信号功率的时间分布。实际上，它就是信道冲激响应中各路径平均功率的分布。假设信道为广义平稳不相关散射信道，即 $h(\tau, t)$ 为广义平稳随机过程，并且信道中第 l 路径的复增益与

第 l '条路径的复增益是不相关的。功率时延分布可表示为

$$P(\tau) = E\{|h(\tau, t)|^2\} = \sum_{l=0}^{L-1} \alpha_l^2 \delta(\tau - \tau_l) \quad (7)$$

式中 $\alpha_l^2 = E[|h_l(t)|^2]$ ，表示第 l 条路径的平均功率。

均方根时延扩展（Root Mean Square delay spread）常用于度量多径信道的时延扩展。实际上，均方根时延扩展即为功率时延分布的两阶中心矩的平方根。

$$\tau_{rms} = \left[\frac{\int_{-\infty}^{\infty} (\tau - \bar{\tau})^2 P(\tau) d\tau}{\int_{-\infty}^{\infty} P(\tau) d\tau} \right]^{1/2} = \left[\frac{\sum_{l=0}^{L-1} \bar{\alpha}_l^2 (\tau_l - \bar{\tau})^2}{\sum_{l=0}^{L-1} \bar{\alpha}_l^2} \right]^{1/2} \quad (8)$$

式中

$$\bar{\tau} = \frac{\int_{-\infty}^{\infty} \tau P(\tau) d\tau}{\int_{-\infty}^{\infty} P(\tau) d\tau} = \frac{\sum_{l=0}^{L-1} \bar{\alpha}_l^2 \tau_l}{\sum_{l=0}^{L-1} \bar{\alpha}_l^2} \quad (9)$$

相干带宽表征一个统计平均带宽，在该带宽内，信道频率特性是相关的。实际上，信道相干带宽与信道均方根时延扩展成反比。相干带宽可定义为信道频率响应自相关函数为某一给定值时对应的频率间隔。例如，频率相关函数大于 0.9 时，对应的相干带宽可近似为

$$B_{coh,0.9} \approx \frac{1}{50\tau_{rms}} \quad (10)$$

更为宽松的定义是频率相关函数大于 0.5 时，对应的相干带宽可近似为

$$B_{coh,0.5} \approx \frac{1}{5\tau_{rms}} \quad (11)$$

信道时间相关函数定义为

$$R_h(\Delta t) = \int_{-\infty}^{+\infty} h(t) h^*(t + \Delta t) dt \quad (12)$$

多普勒频谱即为信道时间相关函数的傅里叶变换

$$S_h(f) = \int_{-\infty}^{+\infty} R_h(\Delta t) e^{-j2\pi f \Delta t} d\Delta t \quad (13)$$

假设发射信号的载频为 f_c ，接收机与发射机的相对移动速度为 v ，多径分量的入射角平均分布于 $[0, 2\pi]$ ，则信道的时延相关函数为

$$R_h(\Delta t) = J_0(2\pi f_c v / c \cdot \Delta t) \quad (14)$$

式中 $J_0(\cdot)$ 为一类零阶贝塞尔函数。相应的多普勒频谱为

$$S_h(f) = \frac{1}{\pi\sqrt{f_d^2 - f^2}}, |f| \leq f_d \quad (15)$$

式中 $f_d = f_c v / c$, 为最大多普勒频移。式(15)即为 Jakes 或 Classical 多普勒频谱。

多普勒扩展和相干时间是描述小尺度区域内信道时变特性的两个参数。在时变信道下, 当传输一个频率为 f_c 的纯正弦信号时, 接收信号的频谱(也就是多普勒频谱)在 $f_c - f_d$ 至 $f_c + f_d$ 范围内存在分量。多普勒扩展就是用于度量由信道时变引起的信号频谱展宽。多普勒扩展可以近似等于最大多普勒频移。当信号带宽远大于多普勒扩展时, 信道为慢衰落信道。

相干时间用于统计度量信道冲击响应在多大的时间间隔内仍具有强相关。当相干时间远远大于码元(符号)周期时, 信道在码元周期内可视为不变, 这时信道为慢衰落信道。而当码元(符号)周期接近相干带宽时, 此时信道为快衰落信道。

若相干时间定义为时间相关函数大于 0.5 时对应的时间间隔, 则相干时间可近似为

$$T_c \approx \frac{9}{16\pi f_d} \quad (16)$$

在现代数字通信系统中, (16)式定义的相干时间过于严格, 一般采用以下定义的相干时间

$$T_c = \sqrt{\frac{9}{16\pi f_d^2}} = \frac{0.423}{f_d} \quad (17)$$

3、信道参数的提取

基于以上信道模型, 提取信道的主要参数有: 多普勒中心频率、多普勒频带宽度、时延扩展等。

为了提取多普勒等频率域参数, 画出对每一场景下的多普勒频谱如下:

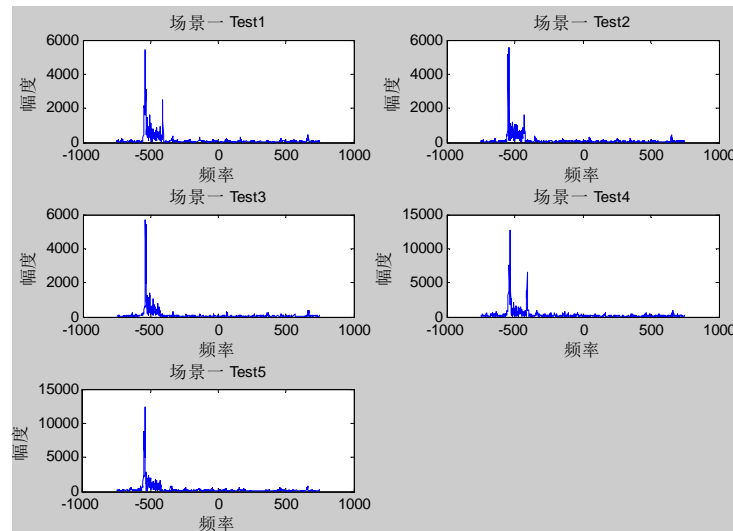


图 4 (a) 场景一的多普勒频谱

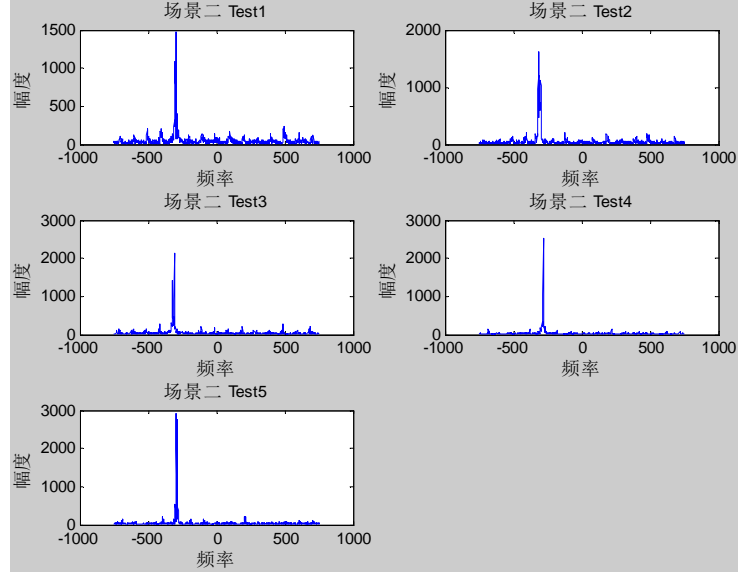


图 4 (b) 场景二的多普勒频谱

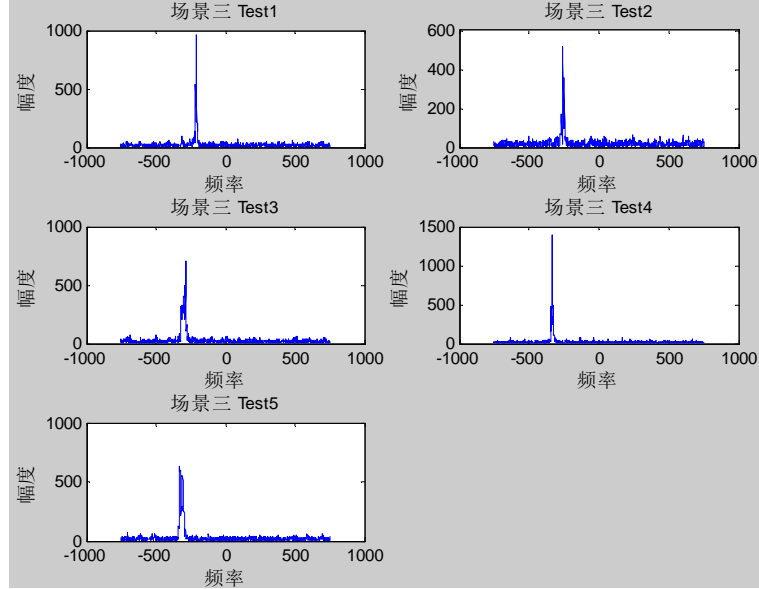


图 4 (c) 场景三的多普勒频谱

从图 4 中可以看出，三个场景中的多普勒频率均为负值，场景一中的多普勒中心频率明显偏离中心频率较大，且多普勒频谱更宽。

多径时延会导致接收端信号的时延扩展，实际测数据为单位脉冲信号经过发射滤波器、信道、接收滤波器后接收到的信号，在没有多径时延和噪声影响的理想情况下，接收信号应为滤波器的单位冲激响应。由于多径时延以及噪声的影响，在接收端收到的信号会被展宽，为了刻画展宽的程度（类似均方根时延扩展），定义时延扩展系数

$$\rho_{\tau} = \frac{\text{信号峰值功率下降90\%的时间宽度}}{\text{滤波器冲击响应的峰值功率下降90\%的时间宽度}}$$

提取得到的三个场景中的多普勒中心频率、多普勒频谱宽度， ρ_{τ} （刻画时

延扩展), 结果下表:

表 2 统计方法提取的信道参数

		多普勒中心 频率 (Hz)	多普勒频谱 宽度 (Hz)	ρ_τ (时延扩 展)
场景一	Test1	-475	155	1.0154
	Test2	-480	160	1.0047
	Test3	-500	160	1.004
	Test4	-490	160	1.0343
	Test5	-550	150	1.0003
	平均	-499	157	1.0117
场景二	Test1	-300	30	1.0004
	Test2	-310	35	1.0003
	Test3	-315	30	1.001
	Test4	-280	20	1.000
	Test5	-290	30	1.000
	平均	-301	29	1.0003
场景三	Test1	-210	30	1.3417
	Test2	-250	35	1.6623
	Test3	-330	30	1.3247
	Test4	-300	43	1.2063
	Test5	-320	44	1.335
	平均	-282	36.4	1.374

从表 2 中数据可以看出, 场景一的多普勒中心频率和频谱宽度明显区别于场景二、三, 但是时延扩展几乎相同; 场景三的时延扩展系数明显大于场景一、二。

4、基于主元数据的特征提取方法:

由于接收信号 $r[k, n]$ 都是二维高阶矩阵, 且含有噪声, 因此直接计算信号信息不仅计算量大, 而且会存在较大误差。如何提取信号的主要分量, 进而对主要分量提取特征是要解决的问题。

为了有效的找出数据的“主要”元素和结构, 去除噪音, 将原有的复杂数据降维, 可以采用多维向量的统计分析方法。

假设以每个样本 x_i 作为列向量形成矩阵 X , 即

$$X = [x_1, x_2, \dots, x_N] \quad (18)$$

寻找一组正交基组成的矩阵 P , 使得

$$Y = PX \quad (19)$$

且 Y 的协方差矩阵

$$C = YY^H \quad (20)$$

为对角阵。将(19)带入(20)得

$$C = PX(PX)^H = PXX^HP \quad (21)$$

令 $A = XX^H$ ，并带入(21)得

$$C = PAP^H \quad (22)$$

对 A 做特征分解得

$$A = EDE^H \quad (23)$$

式中 $E = [p_1 \ p_2 \ \cdots \ p_M]$ 为 A 的特征向量 p_i 组成的矩阵，且有 $E^{-1} = E^H$ ，

$D = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_M)$ 为 A 的特征值组成的对角矩阵。取

$$P = E^H \quad (24)$$

则

$$C = E^H A E = E^H (EDE^H) E = (E^{-1} E) D (E^{-1} E) = D \quad (25)$$

此时， X 的主元也就是 XX^H 的特征向量，也就是矩阵 P 的行向量；矩阵 C 对角线上的第 i 个元素就是数据 X 在方向该对应特征向量的能量，特征值越大该特征向量（即 X 的主元）所占的比重越大。

至此，我们可以用数据样本作为列向量所组成的样本矩阵，对样本矩阵的协方差矩阵做特征分解，求得特征向量以及特征根。样本的特征均体现在特征向量上，而特征值描述了以该特征向量作为数据特征的重要程度。用该方法提取主要特征分量的步骤为：

- 1、选取数据样本，组成样本矩阵 X ；
- 2、对 XX^H 进行特征分解，求取特征向量（即 X 的主元）和特征值；
- 3、选取前 p 个贡献率最大的特征值及其对应的特征向量，贡献率可以

是选取的特征值占有所有特征值之和的比重。

对样本数据选择出主要的特征分量（即主元），就可以对主要的特征分量进行分析，以提取信道的主要的特征。

为了分析信道的时变特性，以测试数据的每一列为样本组成样本矩阵，提取贡献率前 95% 的特征值以及特征向量，对特征向量进行分析。用主元数据提取信道的频率特征如下图

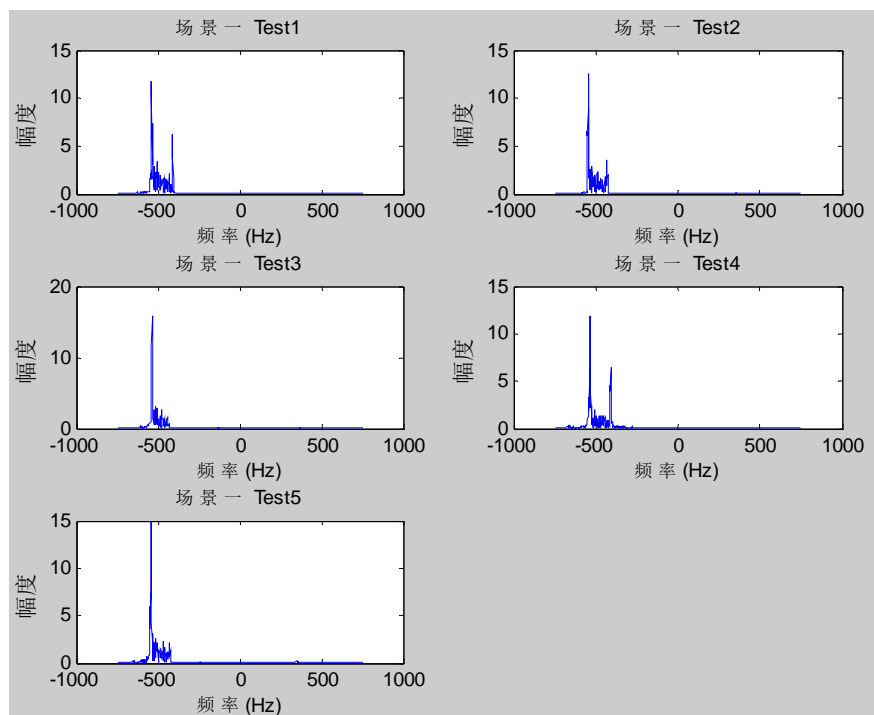


图 5 (a) 场景一主元数据的多普勒频谱

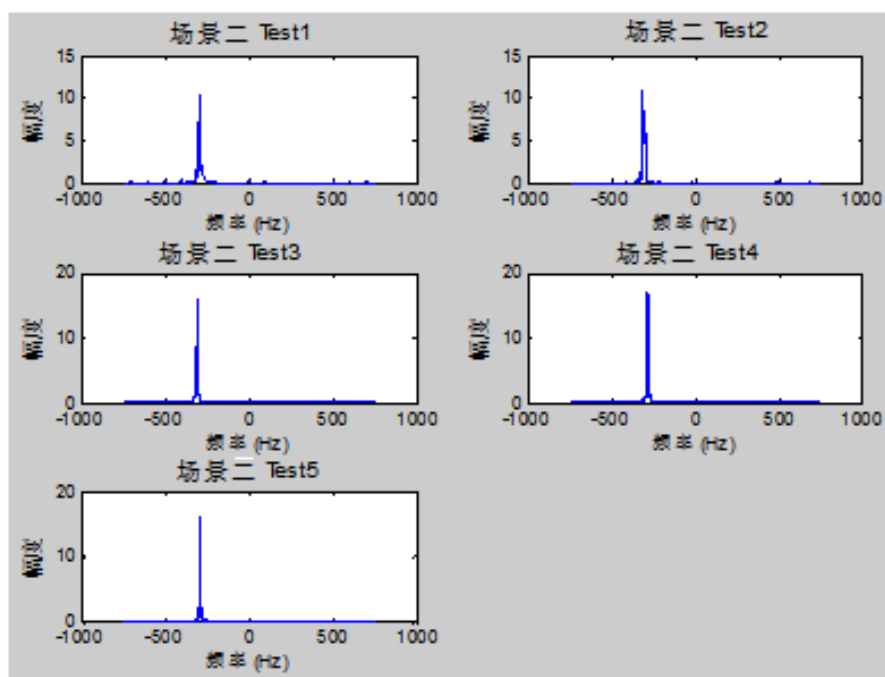


图 5 (b) 场景二主元数据的多普勒频谱

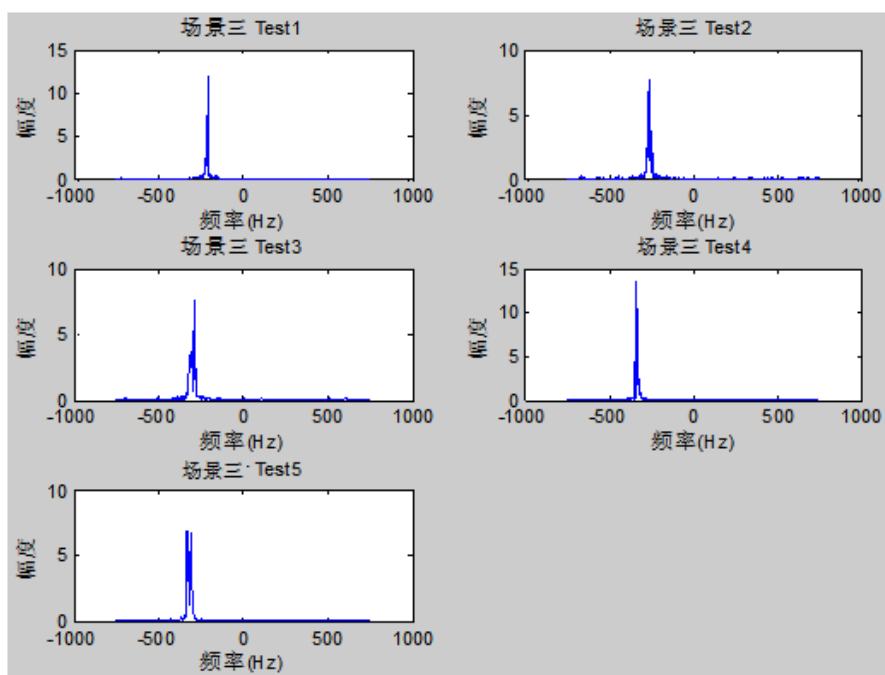


图 5 (c) 场景三主元数据的多普勒频谱

以测试数据的每一行为样本组成样本矩阵,提取贡献率前 95% 的特征值以及特征向量,对特征向量进行分析。用主元数据提取信道的多径延迟数据如下:

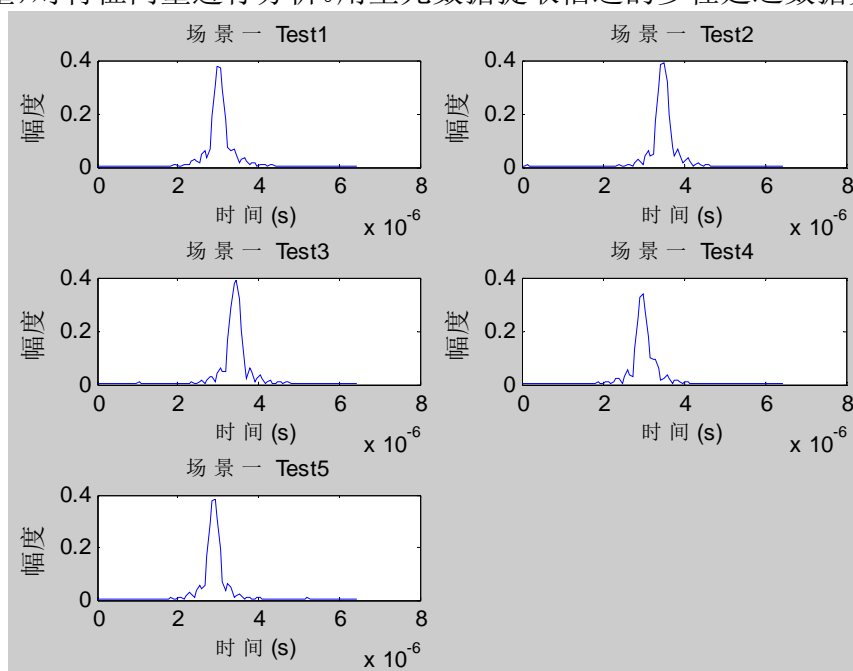


图 6 (a) 场景一主元数据的时域特征

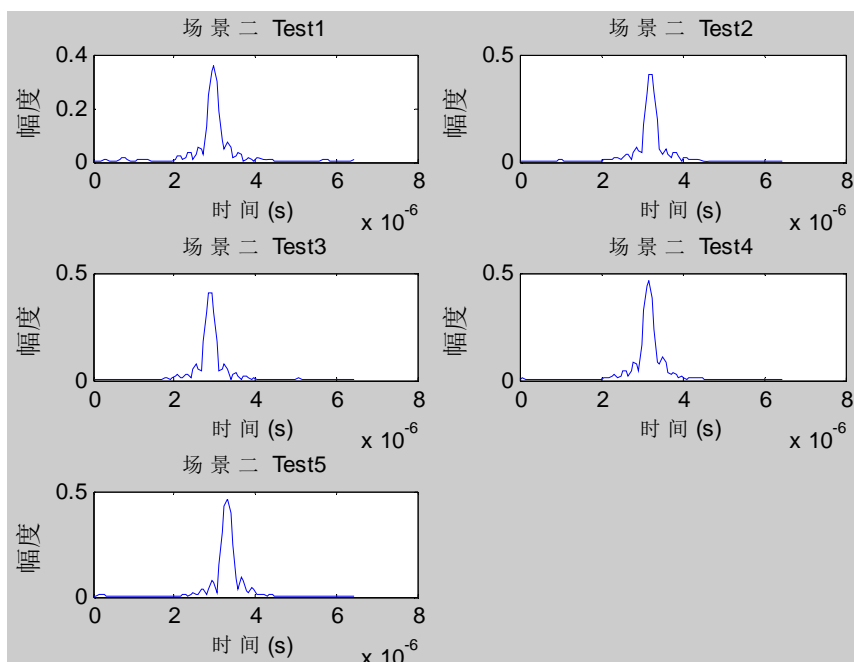


图 6 (b) 场景二主元数据的时域特征

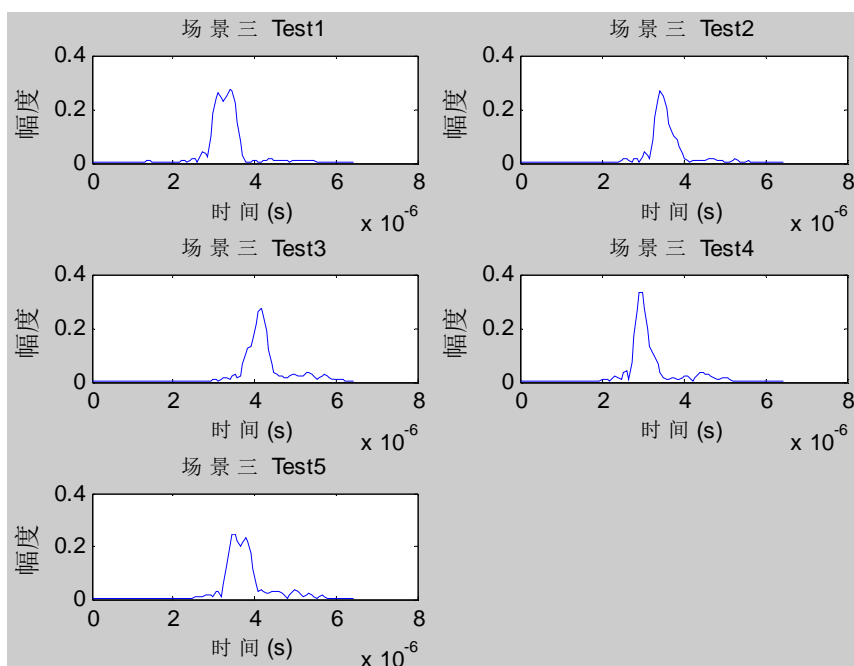


图 6 (c) 场景三主元数据的时域特征

从图中可以发现，提取主元数据，然后对主元数据进行频域和时域特征提取。对比图 4 和图 5，可以看出，两幅图中频谱出现的位置完全一致，并且图 5 中的噪声明显降低；从图 6 可以看出，场景三的信号宽度要明显宽于场景一、二，即说明场景三中信号的时延扩展要大。

从图 4、图 5 和图 6 提取信息如下表：

表 3 基于主元数据的方法提取的信道参数

		多普勒中心频率 (Hz)	多普勒频谱宽度 (Hz)	ρ_τ (时延扩展)
场景一	Test1	-475	155	1
	Test2	-480	160	1
	Test3	-500	160	1
	Test4	-490	160	1
	Test5	-550	150	1
	平均	-499	157	1
场景二	Test1	-300	30	1
	Test2	-310	35	1
	Test3	-315	30	1
	Test4	-280	20	1
	Test5	-290	30	1
	平均	-301	29	1
场景三	Test1	-210	30	1.1429
	Test2	-250	35	1.2857
	Test3	-330	30	1.4286
	Test4	-300	43	1.4286
	Test5	-320	44	1.7134
	平均	-282	36.4	1.4

说明：关于时延扩展系数 ρ_τ ，由于信号能量相对集中，而时间分辨率精度不够，无法给出精确地时延扩展系数的比值，但是从图 6 中可以明显的看出差别。

因此可以得出以下结论，基于主元数据的特征提取可以获得与统计方法一样的效果。基于主元数据的特征提取相比于统计方法的特征提取的优点在于：统计方法的特征提取要对所有的样本的数据进行分析提取特征，当样本很多时计算量是相当大的，而基于主元数据的特征提取只需要对少数几个主元数据进行特征分析，大大减少了分析的运算量。需要指出的是，主元数据的获取仍然是要通过矩阵特征值分解，在变量维数比较多时计算量也是相当大的。

5、信道“指纹”模型的建立

因此，为了建立有效的“指纹”模型，先对提取样本数据的主元数据，然后根据主元数据分析信道的“指纹”，描述“指纹”的参数有多普勒中心频率、多普勒频谱宽度、 ρ_τ （时延扩展），因此建立三维“指纹”模型：

$$Q = \begin{bmatrix} q_1 \\ q_2 \\ q_3 \end{bmatrix} \quad (26)$$

式中 q_1 与多普勒中心频率有关，是由发射端和接收端的相对移动引起的； q_2 与多普勒频谱宽度有关，是由信道的时变特性引起的； q_3 与时延扩展系数有关，是由多径因素引起的。

理论上每一场景都有唯一的“指纹”与之对应，但是“指纹”中不可避免的会包含一些噪声，因此假两个设包含“噪声”的“指纹”可以分别为 Q_i 和 Q_j 。为了评价所指纹所描述场景的准确性，两个场景的差别可以用

$$D = \|Q_i - Q_j\| \quad (27)$$

来描述，考虑到每一个指纹特征代表的信道参数不同，因此更为完善的描述两个场景的差别，可以用如下公式

$$D = \alpha_1 |q_{1,1} - q_{2,1}|^2 + \alpha_2 |q_{1,2} - q_{2,2}|^2 + \alpha_3 |q_{1,3} - q_{2,3}|^2 \quad (28)$$

式中 $q_{i,j}$ 代表第 i 个场景的第 j 个信道参数， $\alpha_1, \alpha_2, \alpha_3$ 分别代表每一信道参数在描述场景差别时所占的比重。由于每个场景总会存在“噪声”，因此完全相同的场景很难存在，可以设立一个适当的阈值 θ ，当

$$D < \theta \quad (29)$$

时，可以认为 Q_i 和 Q_j 是同一个场景。

当影响一个场景的因素增加或减少时，可以对 Q 进行升维或降维处理。比如，当通信系统的接收机与发射机的距离需要考虑时，可以再对 Q 增加一个维度 q_4 来衡量功率，即 $Q = [q_1, q_2, q_3, q_4]^T$ ；再比如，当接收机相对发射机的速度可以不作为描述场景的一个参数时，可以对 Q 降维处理， q_1 代表的多普勒中心频率就可以忽略。

四. 问题二的求解

数据包二中给出了两次实测数据的测量结果，根据问题一中的基于主元特征的信道“指纹”模型，先提取样本数据的主元数据，然后提取“指纹”参数如下

Test1ForScene 的主元数据的频谱图和时域特性图：

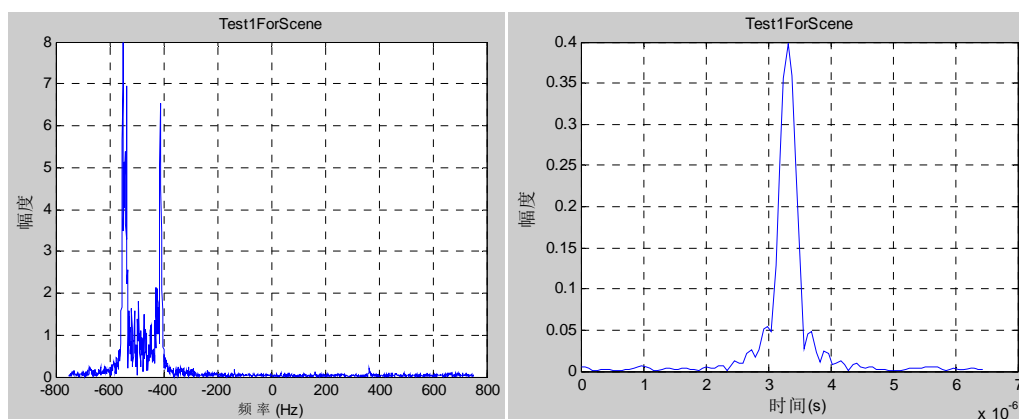


图 7 Test1ForScene 的频域特征和时域特征

Test2ForScene 的主元数据的频谱图和时域特性图:

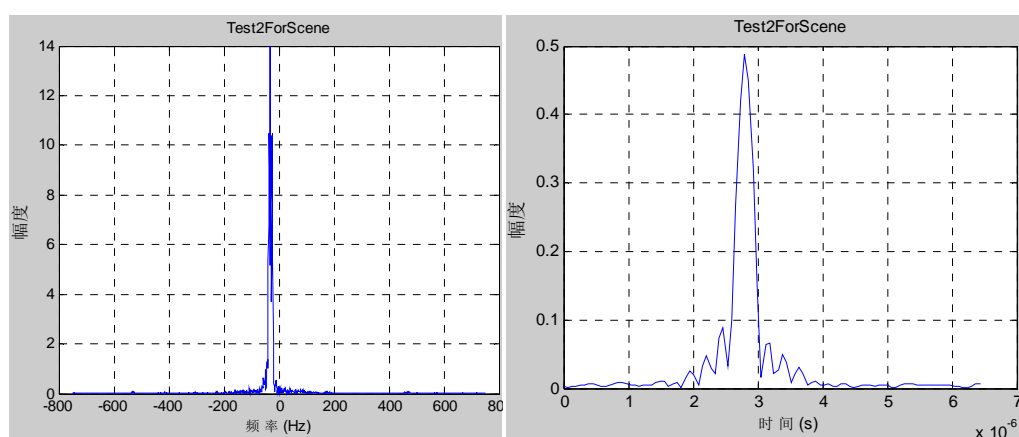


图 8 Test2ForScene 的频域特征和时域特征

将图 7 和图 8 中的信息提取到如下表格中:

	多普勒中心频率 (Hz)	多普勒频谱宽度 (Hz)	时延扩展
Test1ForScene	-480	160	1
Test2ForScene	-30	33	1

对比表 2 和表 3，可以得出结论如下：

一、Test1ForScene 的多普勒中心频率、多普勒频谱宽度、时延扩展均与场景一相似；多普勒中心频率和多普勒频谱宽度两个因素与场景二相差较大；多普勒中心频率、多普勒品谱宽度、时延扩展与场景三均相差较大。因此判定 Test1ForScene 属于场景一。

二、Test2ForScene 的多普勒中心频率、多普勒频谱宽度与场景一差别较大；除多普勒中心频率外，多普勒频谱宽度、时延扩展均与场景二相符；在时延扩展上与场景三相差较大。因此，如果“场景”中不考虑接收端与发射端的相对运动，只描述环境因素时，可以认为 Test2ForScene 属于场景二。

五. 问题三的分析与建模

“数据包 3”提供了一条连续路段的真实信道测量结果。该结果对应于以 3km/h 的速度步行近 150m 的测试距离。该路段可能包含不同的场景或环境，对应不同的“指纹”特征。基于上述测量结果，先自行分段，采用“问题 1”中的建模方法，给出分段的“指纹”分析。在分段分析的基础上，对所提取的“指纹”进行合理分类。通过对比不同段的分析结果，最终决定该路段可以依“指纹”划分为多少个区域。

1、问题分析

为了便于分析，仍然先给出测试数据的等高图，如下图所示

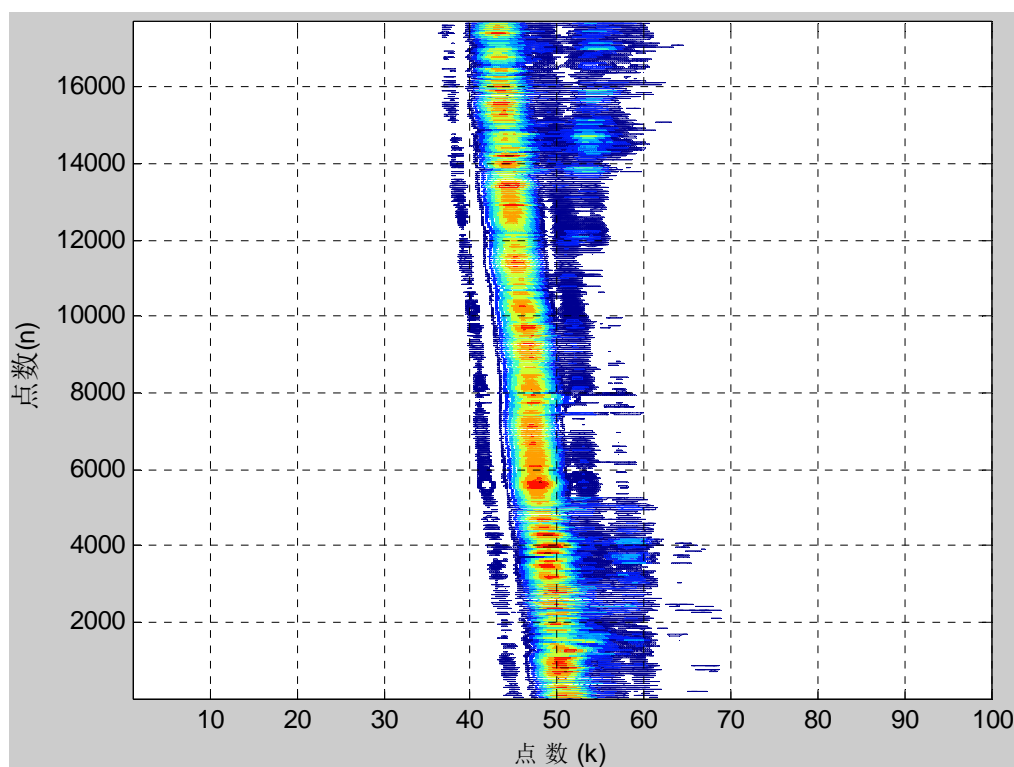


图 9 问题三路段测试数据等高图

从图中可以得到以下分析：

1、随着采样时刻的变化，信号出现峰值的时刻越来越早，说明运动的方向朝着发射源（确切地说，是离发射源越来越近）；

2、通过观察图形可以大概分出以下路段：

（1）采样点 1—5000，在此路段内信号能量集中的时间宽度相对较宽，相比滤波器的冲激响应宽度明显较大，说明此路段内多径效应较为明显；

（2）采样点 5000—7000，在此路段内信号的能量集中的时间宽度与滤波器的冲激响应相比宽度相当，在此路段内，多径效应可能较弱；

（3）采样点 7000—8000，在此路段内信号能量集中在较短的时间范围内，多径效应最弱；

（4）采样点 8000—12000，此路段内信号的能量集中的时间宽度于滤波器的冲激响应相比宽度相当，在此路段内，多径效应也比较弱；

（5）采样点 12000—17700，此路段内随着采样点的增加，信号能量较集中

的时间段出现分离，说明有两条较强的信号路径，随着距离的前进，一条路径的长度在变短，另外一条路径的长度在增加，当此路段内可能存在较强的反射，一条路径上的信号先到达接收机，另外一条路径上较强的信号经过侧方或者后方的建筑物等的反射到达接收机。

2、基于信道“指纹”的区域划分

以上分析只是从时间域上直观地分析，为了用问题一中的“指纹”模型去划分该路段还需要提取更多的参数。下面根据上述分段，采用基于主元数据的特征提取方法提取每一路段的“指纹”，来验证上述分段的合理性。

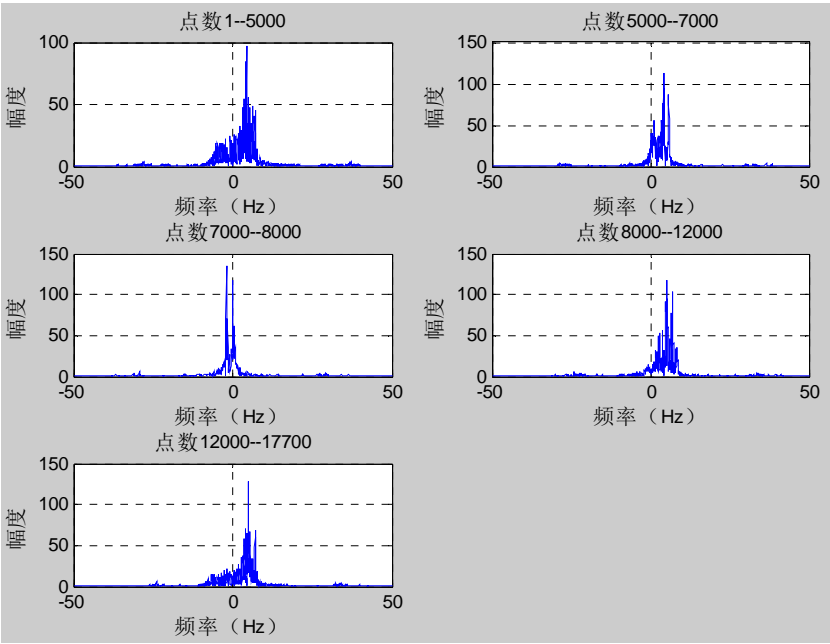


图 9 (a) 不同路段的信道时域特征

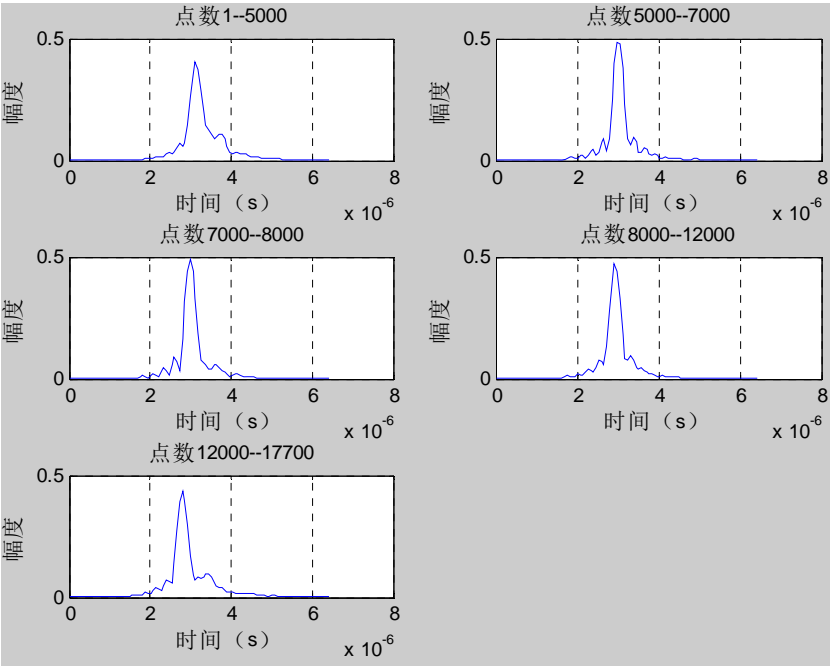


图 9 (b) 不同路段的信道时域特征

为了便于分析数据，将图 9 中的参数提取到以下表格中，由于问题中指出

接收设备是处在移动中的，因此信道的“指纹”可以忽略多普勒中心频率因素，只分析多普勒频谱宽度和时延扩展系数。

	多普勒频谱范围 (Hz)	时延扩展系数
路段一（点数 1-5000）	[-9,8]	1.25
路段二（点数 5000-7000）	[-3,7]	1.04
路段三（点数 7000-8000）	[-5,2]	1.02
路段四（点数 8000-12000）	[-3,9]	1.05
路段五（点数 12000-17700）	[-9,9]	1.30

表中数据表明，通过直观分析划分出来的路段的信道“指纹”确实存在差异，因此可以作为较为粗略的路段划分。将路段的点数根据运动速度和采样时间折算成距离（单位：米）可以看出不同路段信道指纹的区别：

（1）[0,40]为路段一，多普勒频谱很宽表明信道的时变较快，时延扩展很大表明多径效应较为明显；

（2）[40,60]为路段二，多普勒频谱比路段一要窄，表明该路段信道时变相对较弱，时延扩展也不明显表明多径效应也较弱；

（3）[60,67]为路段三，该路段内多普勒频谱和时延效应都最弱；

（4）[67,100]为路段四，该路段内多普勒频谱较宽、时延系数接近于 1，表明该路段内信道时变较强但多径效应不明显；

（5）[100,150]为路段五多普勒频谱很宽、时延系数较大，表明该路段信道时变很强，多径效应也很明显，与路段一有较大相似之处。

更为精细的路段划分可以在粗略划分的基础上再对每一段的部分数据进行分析，从另外一个维度上继续对路段进行划分。理论上，“区域划分”越细，后续做“区域识别”的精确程度越高，但过细的“区域划分”会在“指纹”特征中引入更多的错误，导致误判概率增大。

六. 问题四的求解

“数据包 4”提供了二个真实信道测量结果。基于“问题 3”中的“区域划分”和“问题 1”中的“评价指标”，首先判断此处提供的二个样本是否采集自“问题 3”中所提供的路段。

为了直观地观察两次采集到的数据，画出数据的等高图如图 10，

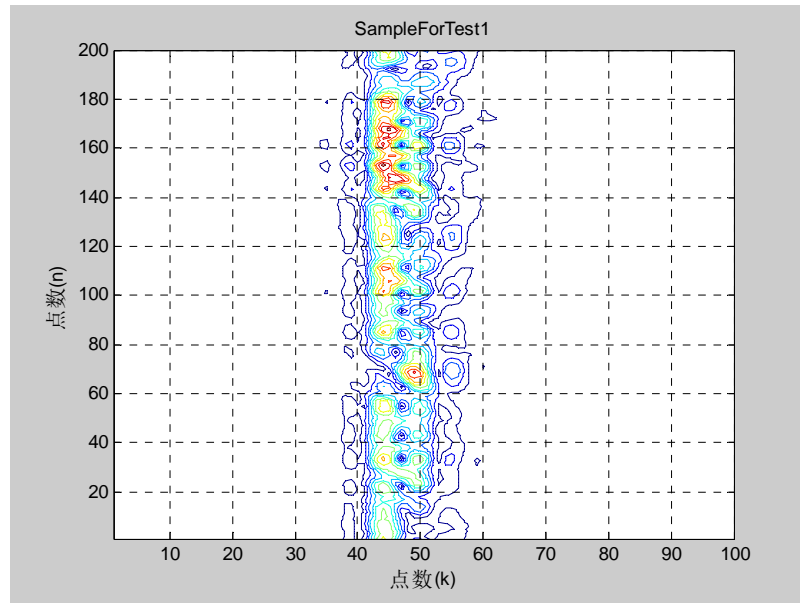


图 10 (a) SampleForTest1 的等高图

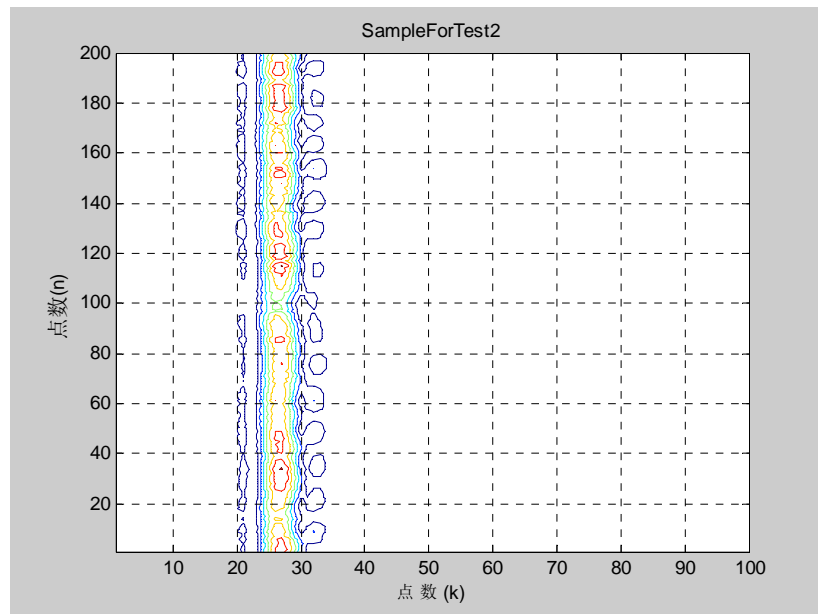


图 10 (b) SampleForTest2 的等高图

从图 10 (a) 中可以看出，采集到的信号能量集中的部分相对较宽，且明显存在两个相对较强的信号，即多径效应较为明显因此猜测 SampleForTest1 可能来自于问题三中的路段一或者路段五；SampleForTest2 的等高图显示在该测试路段，接收到的信号要能量非常集中，可以判断几乎只有一条路径的来波信号，因此猜测该路段可能是位于路段二、路段三或者路段四。为了利用“指纹”特征判断两次测量是否来自于问题三中的路段，提取两次测试数据的“指纹”。

画出 SampleForTest1 的频域和时域特征如图 11 所示，SampleForTest2 的频域和时域特征如图 12 所示

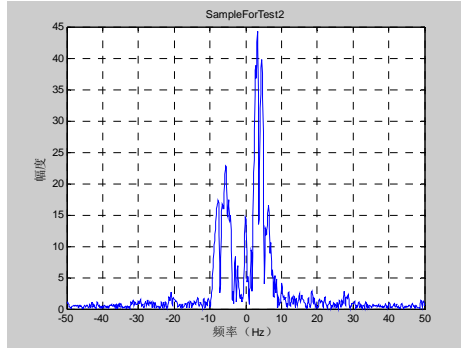


图 11 (a) SampleForTest1 的频谱图

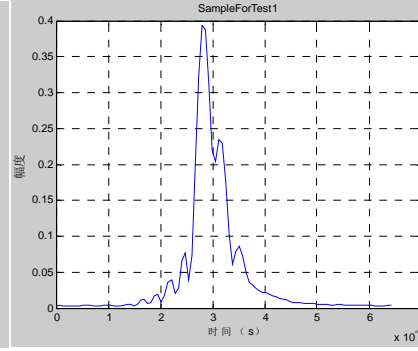


图 11 (b) SampleForTest1 的时域特性

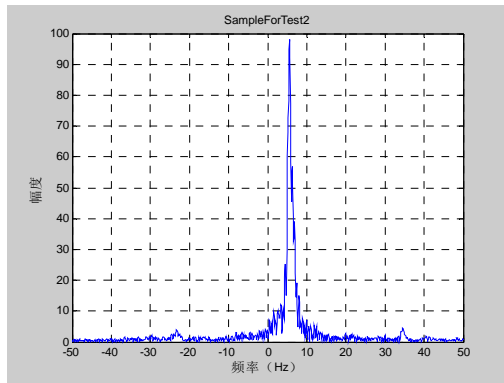


图 12 (a) SampleForTest2 的频谱图

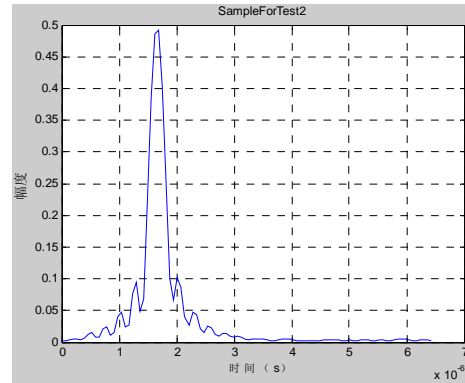


图 12 (b) SampleForTest2 的时域特性

结合图 11 和图 12, SampleForTest1 和 SampleForTest2 的“指纹”下表所示

	多普勒频谱范围 (Hz)	时延扩展系数
SampleForTest1	[-10,10]	1.25
SampleForTest2	[0,10]	1.05

在问题三中路段粗略划分为 5 段的基础上, 对比两次测量场景的指纹和五个路段的指纹进行对比, 可以粗略的认为 SampleForTest1 在路段 1, SampleForTest2 在路段 4。

路段 1 和路段 4 都相对较长, 为了更为精确的判断两次测试的属于的路段。有必要对路段 1 和路段 4 从频谱宽度的维度上进行更为精细的划分。容易想到的方法是对两个路段进行等间隔划分成若干段, 但是划分的路段越多需要分析的数据量就越大, 路段划分的越少, 测试路段的判断就会越模糊, 两者之间需要作出折中的选择。

考虑到粗略划分的路段 1 和路段 4 采样点数分别为 5000 和 4000, 因此选择每一路段的采样点数为 1000 进行更细的划分, 并从频域出发分析不同路段信道时变的差别。

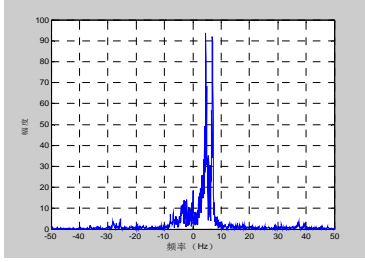


图 13(a) 路段 1 采样点 1-1000

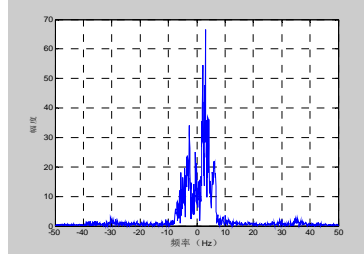


图 13(b) 路段 1 采样点 1000-2000

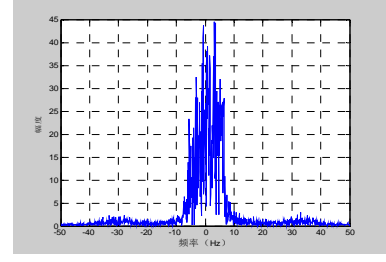


图 13(c) 路段 1 采样点 2000-3000

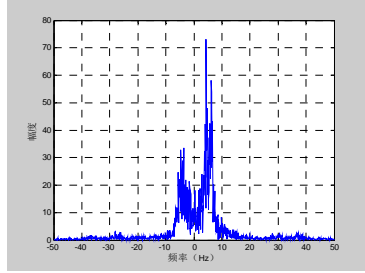


图 13(d) 路段 1 采样点 3000-4000

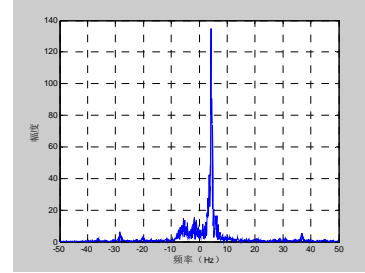


图 13(e) 路段 1 采样点 4000-5000

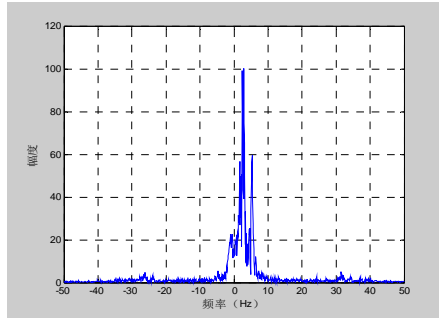


图 14(a) 路段 4 采样点 8000-9000

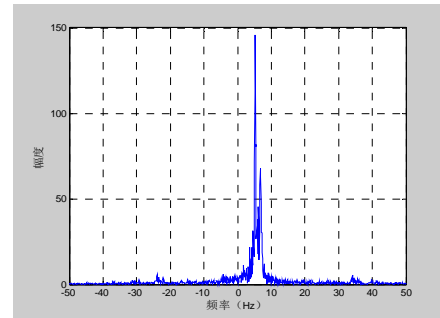


图 14(b) 路段 4 采样点 9000-10000

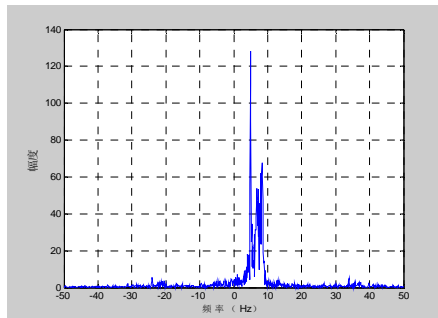


图 14(c) 路段 4 采样点 10000-11000

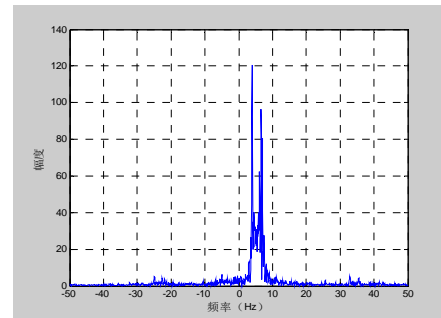


图 14(d) 路段 4 采样点 11000-12000

对比图 11 (a) 和图 13, 图 12 (a) 和图 14, 对频谱宽度做出大概估计, 判断 SampleForTest1 属于路段 1 中的采样点 1000-4000 的概率较大; 判断 SampleForTest2 属于路段 4 中的采样点 9000-11000 的概率较大。

综合以上所有分析, 并折算成距离, 粗略判断 SampleForTest1 来自问题三中路段的 10m—30m, SampleForTest2 来自路段的 75m—90m。

七.模型评价与结论

本文的主要工作是针对无线信道“指纹”特征建模，在先验模型和测试数据的基础上，提取不同场景或不同区域内无线信道的差异化的特征，进而进行分类和归纳。

在对所给数据进行分析后，结合背景资料建立了多径时变信道的复冲激响应模型，以及该模型下的主要参数（时延扩展、多普勒频谱、多普勒扩展）。由于信道的时变具有相关性，因此大量的测试样本之间应该存在较强的相关性，如果通过传统的统计的方法对样本一一分析，分别提取参数，会存在较大的数据运算。本文在 PCA 算法的基础上提出了基于主元数据的信道参数提取的方法，即通过相关矩阵求出特征值与特征向量，并用主要的特征值与特征向量来描述信道样本的，由于矩阵分解完之后只存在为数不多的几个主要特征值，因此用主要特征值与特征向量来代替整个采集样本，进而进行信道参数提取，计算量会明显减小，而且所提取的参数与通过统计的方法获得的参数未见明显差别，效果很好。

利用基于主元数据的方法，提取的信道参数有多普勒频谱、多普勒带宽、时延扩展。通过基于主元数据提取信道参数后，建立了三维的信道“指纹”

$Q=[q_1, q_2, q_3]^T$ ，给出了“指纹”间的“距离” $D=\|Q_1-Q_2\|$ 以及门限 θ 作为评价指标。通过该“指纹”模型可以明显区别出问题中所给的场景，并对场景进行识别；同时，通过指纹的变化可以对连续的场景或区域进行分类。

数据表明，本文提出的基于主元数据的信道参数特征提取在通信信道“指纹”特征建模中拥有较大参考意义。

在信息化的今天，基于大数据的现代数据分析的手段有很多，本文提取的信道参数仅是传统意义上的信道的参数，并有其明确的物理意义。而数据分析技术，比如人工神经网络、支持向量机、聚类分析等在网络数据分析中已广泛应用，能否应用这些数据分析技术分析信道数据，进而提取信道的参数并分析其中的物理意义仍是待研究的问题。由于竞赛时间有限，此项工作将在日后的工作学习中继续进行。

参考文献

- [1] 张弦, 无线通信的信道建模与仿真技术研究[D]:8-9, 2012.
- [2] 郑琛, PCA 算法的原理及其示例[D], 北京师范大学:1-6.
- [3] 宁佐廷, 基于 PCA 的人脸识别算法研究[D], 中南林业大学, 2012
- [4] 徐仲等, 矩阵论简明教程(第四版)[M], 北京, 科学出版社, 2005。
- [5] 张峻铭, 无线信道多径时延估计及信道建模[D], 电子科技大学, 2010
- [6] 易岷, 时延及相关参数估计技术研究[D], 电子科技大学, 2004。
- [7] 李敏, OFDM 系统多径时延及信道估计理论研究和算法实现[D], 兰州理工大学, 2009